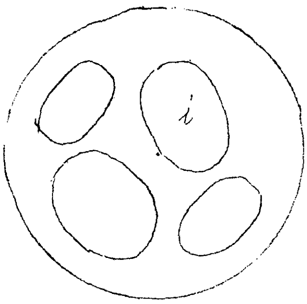
(18) サムプリングに於ける母集団の或る構成方法 ——

林　知己夫

§1. サムプリングによつてある調査対象(Universe)のある量を推定しようとするとき(此の量が唯一種類である場合)通常 Universe の各構成要素に等しい抽出確率をあたへて — 或はことなる抽出確率を與へてある場合があるとしても其は bias をさけるためのものである ——— 母集団を構成し無意識的にサムプリングによる variance を計算して居た。そしてその variance の大小を左右するものは ——— 推定しようとする量の母集団に於ける variance 一定とするとき ——— サムプルの大きさのみであつた。
水野所員の「サムプリング・システム」に関する劃期的論説に於て喝破せられて居るごとくサムプリングによる variance はサムプルの大きさが一定母集団の variance が一定母集団の大きさが一定であつても母集団の構成し方，即ち調査対象の抽出確率のあたへ方によつて小さくなしうるのである。この様にして Universe の調査しようとする量を sharp に推定するためには population を巧に構成することが問題になつてくる。此に関する一般論は前記論説に示されてあるからこゝでは簡単な実例をあげてみよう。

§2. 推定しようとする量は Universe の所謂算術平均とする。 Universe は M 個の Block よりなり各 block



は N 個の element よりなるとする。今 Sampling の system を『i block の抽出される確率は，p_i である。一度 j block が抽出されたとき j block の各 element は等しい確率で抽出される』によつてあたへるものとする。

此の population から一つの Block を抽出それから n 個の Sample を抽出し population の平均を推定することを考へる。この時推定平均の variance を最小にするには P_i を如何にあたへたらよいか。此を問題としてみよう。平均の unbiased estimate として

$$Y_i = \frac{X_1^{(i)} + \cdots + X_n^{(i)}}{P_i \cdot n} \cdot \frac{1}{M} \quad \text{をとる}$$

$X^{(i)}$ は i block 中の抽出された n 個の Sample とする。

$$E(Y_i) = \sum_{i=1}^{M} E\left(\frac{X_1^{(i)} + \cdots + X_n^{(i)}}{n}\right) \cdot \frac{1}{P_i} \times P_i \times \frac{1}{M}$$

$$= \frac{1}{M} \sum_{i=1}^{M} \overline{X}^{(i)} = \overline{X}$$

$\overline{X}^i$, $\overline{X}$ は夫々 i block の平均，population の平均とする．

$$\sigma_s^2 = E(Y - \overline{X})^2 \quad \text{を考へてみる}$$

$$\sigma_s^2 = \sum_i P_i E(Y_i - \overline{X}^i)^2 + \sum P_i (\overline{X}^i - \overline{X})^2$$

$$= \sum_{i=1}^{M} \frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \frac{\sigma_i^2}{P_i M^2} + \sum_{i=1}^{M} P_i \bar{X}_i^2 \left(\frac{1}{P_i M} - 1 \right)^2 + \sum_{i=1}^{M} P_i (\bar{X}_i - \bar{X})^2$$

σ_i^2 は i block の variance とする

さらに見とほしをよくするために簡単な場合を考へ，
$\bar{X}_i = \bar{X}$ とする。

$$\sigma_s^2 = \sum_{i=1}^{M} C \cdot \frac{\sigma_i^2}{P_i M^2} + \sum_{i=1}^{M} P_i \bar{X}^2 \left(\frac{1}{P_i M} - 1 \right)^2$$

但し $C = \frac{N-n}{N-1} \frac{1}{n}$

此の σ_s^2 を $\sum P_i = 1$ の條件の下に minimum にするには P_i を如何に定めたらよいか？

$$f = \sigma_s^2 + \lambda (\sum P_i - 1)$$

$$\frac{\partial f}{\partial P_i} = 0 \qquad i = 1, \text{-------}, M$$

とすれば

$$P_i = \frac{\sqrt{\sigma_i^2 C + \bar{X}^2}}{\sum_{i=1}^{M} \sqrt{\sigma_i^2 C + \bar{X}^2}} \qquad \text{をうる}$$

$$\frac{\sigma_i}{\bar{X}_i} = V_i \quad \text{とおけば}$$

(Γ)

$$P_i = \frac{\sqrt{1 + V_i^2 C}}{\sum_{i=1}^{M} \sqrt{1 + V_i^2 C}} \qquad \text{となる}$$

此が各 block にあたるべき抽出確率となる

§ 3. 例

$$\begin{cases} \sigma_1^2 = 100 \\ \sigma_2^2 = 10 \\ \bar{X}^2 = 10 \end{cases}$$

$$C = \frac{N-1}{N-1}\frac{1}{N} = \frac{1}{10} \quad \text{とする。}$$

$$P_1 = \frac{\sqrt{\ }}{2\sqrt{5}+\sqrt{11}} = 0.58$$

$$P_2 = \frac{\sqrt{11}}{2\sqrt{5}+\sqrt{11}} = 0.42 \quad \text{となる}$$

通常は　$P_1 = P_2 = 0.5$　とおいている。
前者の variance を σ_s^2 後者のを σ^2 とすると

$$\sigma_s^2 \fallingdotseq 4.88$$

$$\sigma^2 = \frac{1}{10}\left(\frac{100+10}{2}\right) = 5.5 \quad \text{となる}$$

$$\frac{\sigma - \sigma_s}{\sigma} = \frac{0.14}{2.34} = 0.06$$

S. D のいみで，約 6 % 精度が上ることとなる。

［注　意］

（P）の式をみるに

$V_i^2 C$ は 1 に比して小さる場合が常である

C は $\underline{\frac{N-n}{N-1}\frac{1}{n}}$

Coefficient variation は通常 1 より小

故に　$V_i^2 C \ll 1$

この時

$$P_i = \frac{1}{M}$$ と通常の場合となる。

§ 4. 次に各 block の大きさが N_i とことなつてゐるとする。この時 §2. と同様に Sampling を行ひ同様の推定をするのであるが，この場合は

$$\dot{Y}^{(i)} = \frac{N_i}{N}\,\frac{x_1^{(i)} + \cdots + x_n^{(i)}}{nP_i}$$ を estimate とする。

$E(Y^i) = \bar{X}$ は明らかである。variance

$E(Y - \bar{x})^2$ を minimum にすることから（同様に $\bar{x}^i = \bar{X}$ とする）— 各 block の抽出確率として —

$$P_i = \frac{N_i\sqrt{1 + C_iV_i^2}}{\sum_{i=1}^{M} N_i\sqrt{1 + C_iV_i^2}}$$

をあたへればよいことを知る。

但し
$$C_i = \frac{N_i - n}{N_i - 1}\,\frac{1}{n}$$

$C_iV_i \ll 1$ ならば

$$P_i = \frac{N_i}{\sum N_i}$$ となる通常の場合の Size Proportionate の確率で抽出する Sampling に帰着する。